MATHEMATICAL ENGINEERING TECHNICAL REPORTS

Improved Loss Estimation for a Normal Mean Matrix

Takeru MATSUDA and William E. STRAWDERMAN

METR 2017–12 April 2017

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html

Improved Loss Estimation for a Normal Mean Matrix

Takeru MATSUDA[1] and William E. STRAWDERMAN[2]

[1]Department of Mathematical Informatics
Graduate School of Information Science and Technology
The University of Tokyo
matsuda@mist.i.u-tokyo.ac.jp
[2]Department of Statistics and Biostatistics
Rutgers University

April 13, 2017

Abstract

We investigate loss estimation in the matrix mean estimation problem. Specifically, for estimators of a normal mean matrix, we consider the estimation of the Frobenius loss. Based on the singular values of the observation, we develop loss estimators that dominate the unbiased loss estimator for a broad class of matrix mean estimators including the Efron–Morris estimator. This is an extension of the results of Johnstone (1988) for a normal mean vector. We also provide improved estimators of loss for reduced-rank estimators. Numerical results show the effectiveness of the proposed loss estimators.

1 Introduction

This paper considers loss estimation for a normal mean matrix. Specifically, let $X \in \mathbb{R}^{n \times m}$ be a matrix observation whose entries X_{ij} are independent normal random variables with mean M_{ij} and variance 1, respectively. In the notation of matrix-variate normal distributions by Dawid (1981), it is expressed as $X \sim \mathrm{N}_{n,m}(M, I_n, I_m)$, where I_k denotes the k-dimensional identity matrix. We assume $n \geq m$. The matrix $M = \mathrm{E}[X] \in \mathbb{R}^{n \times m}$ is called the mean matrix. We consider estimation of M under the Frobenius loss:

$$L(M, \hat{M}(X)) = \| \hat{M}(X) - M \|_{\mathrm{F}}^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} (\hat{M}_{ij}(X) - M_{ij})^2. \tag{1}$$

Note that the Frobenius loss (1) depends on both M and X. Let $\lambda(X)$ be an estimator of the Frobenius loss (1). We evaluate the performance of a loss estimator $\lambda(X)$ by squared error:

$$L^*(M, \hat{M}(X), \lambda(X)) = (\lambda(X) - L(M, \hat{M}(X)))^2.$$

A loss estimator $\lambda_1(X)$ is said to dominate (or improve) another loss estimator $\lambda_2(X)$ if

$$\mathrm{E}_M[L^*(M, \hat{M}(X), \lambda_1(X))] \leq \mathrm{E}_M[L^*(M, \hat{M}(X), \lambda_2(X))]$$

holds for every M, with strict inequality for at least one value of M. One standard loss estimator in normal mean estimation problems is Stein's unbiased risk estimator (Stein, 1974). This estimator is the unique unbiased loss estimator. The main focus of this paper is on developing loss estimators that dominate the unbiased loss estimator for a broad class of matrix mean estimators.

There has been substantial work on loss estimation for a normal mean vector, which corresponds to $m = 1$. For the maximum likelihood estimator, Johnstone (1988) showed that the unbiased loss estimator is improved by negative correction when $n \geq 5$. Johnstone (1988) also showed that, for the James–Stein estimator, the unbiased loss estimator is improved by positive correction when $n \geq 5$. However, loss estimation for a normal mean matrix ($m > 1$) has not been investigated.

In this paper, we focus on the class of equivariant estimators for a normal mean matrix:

$$\hat{M}(X) = U\Sigma(I_m - \Phi(\Sigma))V^\top, \tag{2}$$

where $X = U\Sigma V^\top$ with $U \in \mathbb{R}^{n\times m}$, $V \in \mathbb{R}^{m\times m}$, $U^\top U = V^\top V = I_m$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ is the singular value decomposition of X and $\Phi(\Sigma) = \mathrm{diag}(\phi_1(\sigma), \cdots, \phi_m(\sigma))$ with $\sigma = (\sigma_1, \cdots, \sigma_m)$. Many matrix mean estimators belong to the above class. For example, Efron and Morris (1972) proposed a minimax estimator that is a natural extension of the James–Stein estimator:

$$\hat{M}_{\mathrm{EM}} = X\left(I_m - (n-m-1)(X^\top X)^{-1}\right),$$

which is rewritten as the form (2) with

$$\phi_i(\sigma) = \frac{n-m-1}{\sigma_i^2} \quad (i = 1, \ldots, m).$$

Minimax estimators of the form (2) have been further developed (Efron and Morris, 1976; Stein, 1974; Zheng, 1986; Tsukuma, 2008). In addition, the reduced-rank estimator, which truncates lower singular values of X, has the form (2) with

$$\phi_i(\sigma) = \begin{cases} 0 & (1 \leq i \leq k) \\ 1 & (k+1 \leq i \leq m) \end{cases}.$$

This estimator is effective in multivariate linear regression since the regression coefficient matrix often has low rank in practice (Reinsel and Velu, 1998). We develop loss estimators that dominate the unbiased loss estimator for mean matrix estimators of the form (2). Our results include an extension of the results by Johnstone (1988) for a normal mean vector.

In section 2, we review existing results on loss estimation for a normal mean vector and matrix mean estimation. In section 3, we develop a formula of the unbiased loss estimators for a normal mean matrix, following the method of Sheena (1995) for the covariance estimation problem. In section 4, we develop improved loss estimators for a broad class of matrix mean estimators, including the Efron–Morris estimator and reduced-rank estimators. In section 5, we show the performance of improved loss estimators by numerical experiments. In section 6, we give some concluding remarks.

2 Preliminaries

2.1 Loss estimation for a normal mean vector

Suppose that we estimate $\theta \in \mathbb{R}^p$ from an observation $X \sim \mathrm{N}_p(\theta, I_p)$ by an estimator $\hat{\theta}(x) = x + g(x)$, where I_k denotes the k-dimensional identity matrix. We are interested in estimation of the quadratic loss $L(\theta, \hat{\theta}(x)) = \|\hat{\theta}(x) - \theta\|^2$, which depends on both θ and x. This problem is called loss estimation (Johnstone, 1988) and has been investigated by several studies (Lu and Berger, 1989; Fourdrinier and Wells, 2012). The performance of a loss estimator $\lambda(x)$ is evaluated by squared error $L^*(\theta, \hat{\theta}(x), \lambda(x)) = (\lambda(x) - L(\theta, \hat{\theta}(x)))^2$. A loss estimator $\lambda_1(x)$ is said to dominate another loss estimator $\lambda_2(x)$ if

$$\mathrm{E}_\theta[L^*(\theta, \hat{\theta}(x), \lambda_1(x))] \leq \mathrm{E}_\theta[L^*(\theta, \hat{\theta}(x), \lambda_2(x))]$$

holds for every θ, with strict inequality for at least one value of θ. A loss estimator $\lambda(x)$ is said to be admissible if any loss estimator does not dominate $\lambda(x)$ and a loss estimator $\lambda(x)$ is said to be inadmissible if there exists a loss estimator that dominates $\lambda(x)$.

The quadratic risk of the estimator $\hat{\theta}$ is defined as $R(\theta, \hat{\theta}) = \mathrm{E}_\theta[L(\theta, \hat{\theta}(x))]$. Stein (1974) showed that $R(\theta, \hat{\theta}) = \mathrm{E}_\theta[\lambda^{\mathrm{U}}(x)]$, where $\lambda^{\mathrm{U}}(x) = p + 2\nabla \cdot g(x) + \|g(x)\|^2$ is called Stein's unbiased risk estimator (SURE). By completeness, $\lambda^{\mathrm{U}}(x)$ is the unique unbiased loss estimator. Johnstone (1988) showed that Stein's unbiased risk estimator is inadmissible for the maximum likelihood estimator and also the James–Stein estimator as follows. We assume that all expectations are finite.

Lemma 1. *(Johnstone, 1988) Consider an estimator $\hat{\theta}(x) = x + g(x)$. Let $h : \mathbb{R}^p \to \mathbb{R}$ be a twice weakly differentiable function. The risk difference between the loss* *estimator* $\lambda(x) = \lambda^{\mathrm{U}}(x) + h(x)$ *and the unbiased loss estimator* $\lambda^{\mathrm{U}}(x)$ *is expressed as*

$$\mathrm{E}_\theta[L^*(\theta, \hat{\theta}(x), \lambda(x))] - \mathrm{E}_\theta[L^*(\theta, \hat{\theta}(x), \lambda^{\mathrm{U}}(x))] = \mathrm{E}_\theta[-2\Delta h - 4g^\top \nabla h + h^2].$$

Therefore, if h *satisfies*

$$-2\Delta h - 4g^\top \nabla h + h^2 \leq 0$$

for every x*, then* $\lambda(x)$ *dominates* $\lambda^{\mathrm{U}}(x)$*.*

Lemma 2. *(Johnstone, 1988) Consider the maximum likelihood estimator* $\hat{\theta}(x) = x$*. If* $p \geq 5$ *and* $0 < r < 4(p-4)$*, then the loss estimator*

$$\lambda(x) = \lambda^{\mathrm{U}}(x) - \frac{r}{\|x\|^2}$$

dominates the unbiased loss estimator $\lambda^{\mathrm{U}}(x) = p$*.*

Lemma 3. *(Johnstone, 1988) Consider the James–Stein estimator* $\hat{\theta}(x) = (1 - \frac{p-2}{\|x\|^2})x$*. If* $p \geq 5$ *and* $0 < r < 4p$*, then the loss estimator*

$$\lambda(x) = \lambda^{\mathrm{U}}(x) + \frac{r}{\|x\|^2}$$

dominates the unbiased loss estimator $\lambda^{\mathrm{U}}(x) = p - \frac{(p-2)^2}{\|x\|^2}$*.*

Fourdrinier and Strawderman (2003) developed a general method for obtaining an improved loss estimator for pseudo-Bayes estimators. Recently, Narayanan and Wells (2015) discussed loss estimation for the LASSO. See Fourdrinier and Wells (2012) for a general review of results on the topic of loss estimation.

2.2 Estimation of a normal mean matrix

Suppose that we have a matrix observation $X \in \mathbb{R}^{n \times m}$ whose entries are independent normal random variables $X_{ij} \sim \mathrm{N}(M_{ij}, 1)$, where $n \geq m$ and $M \in \mathbb{R}^{n \times m}$ is an unknown mean matrix. In the notation of matrix-variate normal distributions by Dawid (1981), it is expressed as $X \sim \mathrm{N}_{n,m}(M, I_n, I_m)$. We consider estimation of M under the Frobenius loss:

$$l(M, \hat{M}(X)) = \|\hat{M}(X) - M\|_{\mathrm{F}}^2 = \sum_{i=1}^n \sum_{j=1}^m (\hat{M}_{ij}(X) - M_{ij})^2.$$

Let $X = U \Sigma V^\top$, $U \in \mathbb{R}^{n \times m}$, $V \in \mathbb{R}^{m \times m}$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ be the singular value decomposition of X, where $U^\top U = V^\top V = I_m$ and $\sigma_1 \geq \cdots \geq \sigma_m \geq 0$ are the singular values of X. In the following, we consider estimators of the form

$$\hat{M} = U \Sigma (I_m - \Phi(\Sigma)) V^\top, \tag{3}$$

where $\Phi(\Sigma) = \mathrm{diag}(\phi_1(\sigma), \cdots, \phi_m(\sigma))$ and $\sigma = (\sigma_1, \cdots, \sigma_m)$. These estimators are equivariant with respect to the transformation $X \to PXQ$ and $M \to PMQ$ with orthogonal matrices $P \in O(n)$ and $Q \in O(m)$. The maximum likelihood estimator $\hat{M} = X$ has the form (3) with $\phi_1(\sigma) = \cdots = \phi_m(\sigma) = 0$. We note that the estimator (3) is expressed as $\hat{M} = X + g(X)$ with $g(X) = -U\Sigma\Phi(\Sigma)V^\top$.

Many minimax estimators of the form (3) have been proposed when $n \geq m+2$. For example, the estimator by Efron and Morris (1972)

$$\hat{M}_{\mathrm{EM}} = X\left(I_m - (n-m-1)(X^\top X)^{-1}\right) \tag{4}$$

corresponds to

$$\phi_i(\sigma) = \frac{n-m-1}{\sigma_i^2} \quad (i = 1, \ldots, m).$$

$\hat{M}_{\mathrm{EM}}$ is minimax. Later, Efron and Morris (1976) proposed the modified estimator

$$\hat{M}_{\mathrm{MEM}} = X\left(I_m - (n-m-1)(X^\top X)^{-1} - \frac{m^2+m-2}{\mathrm{tr}(X^\top X)} I_m\right), \tag{5}$$

which corresponds to

$$\phi_i(\sigma) = \frac{n-m-1}{\sigma_i^2} + \frac{m^2+m-2}{\sum_j \sigma_j^2} \quad (i = 1, \ldots, m).$$

$\hat{M}_{\mathrm{MEM}}$ is minimax and dominates $\hat{M}_{\mathrm{EM}}$. Zheng (1986) also developed minimax estimators of the form (3) including an extension of the Baranchik type estimators.

Bayes minimax estimators of the form (3) have also been developed. Tsukuma (2008) proposed a class of hierarchical priors and proved admissibility and minimaxity of the Bayes estimators based on them. His prior is a natural generalization of Strawderman's prior. On the other hand, Matsuda and Komaki (2015) developed a singular value shrinkage prior

$$\pi_{\mathrm{SVS}}(M) = \det(M^\top M)^{-(n-m-1)/2}, \tag{6}$$

which is superharmonic. This prior is a natural generalization of Stein's prior. The Bayes estimator with respect to π_{SVS} is minimax and has similar properties to $\hat{M}_{\mathrm{EM}}$. This is an extension of the relationship between the James–Stein estimator and Stein's prior.

Reduced-rank estimators, which are commonly used in multivariate linear regression, also has the form (3). When we apply multivariate linear regression to real data, the regression coefficient matrix often has low rank. Therefore, it is effective to impose a rank constraint on the regression coefficient matrix and this method is called the reduced-rank regression (Reinsel and Velu, 1998). Here, we consider the reduced-rank estimator with rank k, which is represented as (3) with

$$\phi_i(\sigma) = \begin{cases} 0 & (1 \leq i \leq k) \\ 1 & (k+1 \leq i \leq m) \end{cases}. \tag{7}$$

Thus, the reduced-rank estimator truncates lower singular values. From the Eckart–Young theorem, this estimator is the maximum likelihood estimator under the rank constraint rank $M = k$ (Eckart and Young, 1936).

3 Unbiased loss estimator for a normal mean matrix

In this section, we derive the unbiased loss estimators for matrix mean estimators of the form (3). Our derivation is similar to the method of Sheena (1995), which provides the unbiased estimator of risk for orthogonally invariant estimators of a covariance matrix. Our method is valid for estimators with discontinuities such as reduced-rank estimators. Takemura and Kuriki (1999) used a similar technique.

Following Sheena (1995), we utilize the coordinate system on the space of matrices derived from the singular value decomposition. For $X \in \mathbb{R}^{n \times m}$, let $X = U\Sigma V^\top$ with $U \in \mathbb{R}^{n \times m}$, $V \in \mathbb{R}^{m \times m}$, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ be the singular value decomposition of X, where $U^\top U = V^\top V = I_m$ and $\sigma_1 \geq \cdots \geq \sigma_m \geq 0$ are the singular values of X. We put $\sigma = (\sigma_1, \cdots, \sigma_m)$, $U = (u_1, \cdots, u_m)$, and $V = (v_1, \cdots, v_m)$, where $u_i \in \mathbb{R}^n$ and $v_i \in \mathbb{R}^m$ are column vectors. From James (1954), the Jacobian of the singular value decomposition is

$$\mathrm{d}X = \prod_{i=1}^m \sigma_i^{n-m} \prod_{i<j} (\sigma_i^2 - \sigma_j^2) \mathrm{d}U \mathrm{d}\Sigma \mathrm{d}V,$$

where

$$\mathrm{d}U = \prod_{i,j=1}^m u_j^\top \mathrm{d}u_i, \quad \mathrm{d}\Sigma = \prod_{i=1}^m \mathrm{d}\sigma_i, \quad \mathrm{d}V = \prod_{i<j} v_j^\top \mathrm{d}v_i.$$

Here, $\prod_{i<j}$ denotes $\prod_{i=1}^m \prod_{j=i+1}^m$. Note that $\mathrm{d}U$ and $\mathrm{d}V$ are the invariant measures on the Stiefel manifold $V(n, m)$ and the orthogonal group $O(m)$, respectively. Thus, when $X \sim \mathrm{N}_{n,m}(M, I_n, I_m)$, the probability density of U, Σ, V with respect to $\mathrm{d}U \mathrm{d}\Sigma \mathrm{d}V$ is

$$\begin{aligned} p(U, \Sigma, V) &= (2\pi)^{-nm/2} \exp\left(-\frac{1}{2}\|X - M\|_\mathrm{F}^2\right) \prod_{i=1}^m \sigma_i^{n-m} \prod_{i<j} (\sigma_i^2 - \sigma_j^2) \\ &= C \exp\left(-\frac{1}{2}\sum_{i=1}^m \sigma_i^2 + \sum_{i=1}^m a_i \sigma_i\right) \prod_{i=1}^m \sigma_i^{n-m} \prod_{i<j} (\sigma_i^2 - \sigma_j^2), \end{aligned}$$

where $C = (2\pi)^{-nm/2}\exp(-\|M\|_{\mathrm{F}}^2/2)$ and $a_i = (U^\top MV)_{ii}$. Therefore, for any statistic $T(\sigma)$ that has finite expectation $\mathrm{E}_M[T(\sigma)] < \infty$, we have

$$\mathrm{E}_M[T(\sigma)] = \int T(\sigma)F(\sigma)G(\sigma)\mathrm{d}\Sigma, \tag{8}$$

where

$$F(\sigma) = C\exp\left(-\frac{1}{2}\sum_{i=1}^m \sigma_i^2\right)\prod_{i=1}^m \sigma_i^{n-m}\prod_{i<j}(\sigma_i^2-\sigma_j^2),$$

and

$$G(\sigma) = \int \exp\left(\sum_{i=1}^m a_i\sigma_i\right)\mathrm{d}U\mathrm{d}V.$$

We note that

$$\frac{\partial}{\partial\sigma_i}G(\sigma) = \int a_i\exp\left(\sum_{i=1}^m a_i\sigma_i\right)\mathrm{d}U\mathrm{d}V. \tag{9}$$

Theorem 1. *Suppose that an estimator $\hat{M}$ of the form* (3) *satisfies the following conditions:*

1. *All expectations in* (10) *are finite.*
2. *For $1 \le i \le m$, $\phi_i(\sigma)$ is absolutely continuous with respect to σ_i.*
3. *For $1 \le i \le m$, $\lim_{\sigma_i\to\sigma_{i-1}}\phi_i(\sigma)F(\sigma)G(\sigma) = \lim_{\sigma_i\to\sigma_{i+1}}\phi_i(\sigma)F(\sigma)G(\sigma) = 0$.*

Then,

$$\mathrm{E}_M[\|\hat{M}-M\|_{\mathrm{F}}^2] = nm+\mathrm{E}_M\left[\sum_{i=1}^m\left(\sigma_i^2\phi_i^2-2(n-m+1)\phi_i-2\sigma_i\frac{\partial\phi_i}{\partial\sigma_i}\right)-4\sum_{i<j}\frac{\sigma_i^2\phi_i-\sigma_j^2\phi_j}{\sigma_i^2-\sigma_j^2}\right]. \tag{10}$$

Proof. Let $g(X) = -U\Phi(\Sigma)V^\top$ so that $\hat{M} = X + g(X)$. Then, the Frobenius risk of the estimator $\hat{M}$ is

$$\begin{aligned}
\mathrm{E}_M[\|\hat{M}-M\|_{\mathrm{F}}^2] &= \mathrm{E}_M[\|X-M\|_{\mathrm{F}}^2] + 2\mathrm{E}_M[\mathrm{tr}(X-M)^\top g(X)] + \mathrm{E}_M[\|g(X)\|_{\mathrm{F}}^2] \\
&= nm + 2\mathrm{E}_M[\mathrm{tr}X^\top g(X)] - 2\mathrm{E}_M[\mathrm{tr}M^\top g(X)] + \mathrm{E}_M\left[\sum_{i=1}^m \phi_i(\sigma)^2\right] \\
&= nm - 2\mathrm{E}_M[\mathrm{tr}\Sigma\Phi(\Sigma)] - 2\mathrm{E}_M[\mathrm{tr}U^\top MV\Phi(\Sigma)] + \mathrm{E}_M\left[\sum_{i=1}^m \phi_i(\sigma)^2\right] \\
&= nm + \mathrm{E}_M\left[\sum_{i=1}^m(\phi_i(\sigma)-2\sigma_i-2a_i)\phi_i(\sigma)\right].
\end{aligned} \tag{11}$$

Let $\Sigma_i = \{(\sigma_1, \cdots, \sigma_{i-1}, \sigma_{i+1}, \cdots, \sigma_m) | \sigma_1 > \cdots > \sigma_{i-1} > \sigma_{i+1} > \cdots > \sigma_m\}$ and $\mathrm{d}\Sigma_i = \mathrm{d}\sigma_1 \cdots \mathrm{d}\sigma_{i-1} \mathrm{d}\sigma_{i+1} \cdots \mathrm{d}\sigma_m$ for $i = 1, \cdots, m$. From (8), (9), and the conditions 2 and 3,

$$\begin{aligned}
\mathrm{E}_M \left[\sum_{i=1}^m a_i \phi_i(\sigma) \right] &= \sum_{i=1}^m \int \phi_i(\sigma) F(\sigma) \frac{\partial}{\partial \sigma_i} G(\sigma) \mathrm{d}\Sigma \\
&= \sum_{i=1}^m \int_{\Sigma_i} \int_{\sigma_{i+1}}^{\sigma_{i-1}} \phi_i(\sigma) F(\sigma) \frac{\partial}{\partial \sigma_i} G(\sigma) \mathrm{d}\sigma_i \mathrm{d}\Sigma_i \\
&= \sum_{i=1}^m \int_{\Sigma_i} \left([\phi_i(\sigma) F(\sigma) G(\sigma)]_{\sigma_i = \sigma_{i+1}}^{\sigma_i = \sigma_{i-1}} - \int_{\sigma_{i+1}}^{\sigma_{i-1}} \frac{\partial}{\partial \sigma_i} \left(\phi_i(\sigma) F(\sigma) \right) G(\sigma) \mathrm{d}\sigma_i \right) \mathrm{d}\Sigma_i \\
&= - \sum_{i=1}^m \int_{\Sigma_i} \int_{\sigma_{i+1}}^{\sigma_{i-1}} \frac{1}{F(\sigma)} \frac{\partial}{\partial \sigma_i} \left(\phi_i(\sigma) F(\sigma) \right) F(\sigma) G(\sigma) \mathrm{d}\sigma_i \mathrm{d}\Sigma_i \\
&= - \sum_{i=1}^m \mathrm{E}_M \left[\frac{1}{F(\sigma)} \frac{\partial}{\partial \sigma_i} \left(\phi_i(\sigma) F(\sigma) \right) \right]. \qquad (12)
\end{aligned}$$

Substituting (12) into (11) and calculating the derivative, we obtain (10). □

We note that all estimators appearing in this paper satisfy the conditions of Theorem 1. From Theorem 1, the unbiased loss estimator is obtained as follows.

Corollary 1. *Suppose that an estimator $\hat{M}$ of the form* (3) *satisfies the conditions of Theorem 1. Then, the unbiased loss estimator for $\hat{M}$ is*

$$\lambda^{\mathrm{U}}(X) = nm + \sum_{i=1}^m \left(\sigma_i^2 \phi_i^2 - 2(n-m+1)\phi_i - 2\sigma_i \frac{\partial \phi_i}{\partial \sigma_i} \right) - 4 \sum_{i<j} \frac{\sigma_i^2 \phi_i - \sigma_j^2 \phi_j}{\sigma_i^2 - \sigma_j^2}. \quad (13)$$

4 Improving on the unbiased loss estimator for a normal mean matrix

In this section, we provide loss estimators that dominate the unbiased loss estimator (13) for matrix mean estimators of the form (3). Our results include an extension of the results by Johnstone (1988) for a normal mean vector.

Throughout this section, we use the following formula from Stein (1974) for the Laplacian form. Note that $\sum_{i<j}$ denotes $\sum_{i=1}^m \sum_{j=i+1}^m$.

Lemma 4. *(Stein, 1974) Suppose that $f : \mathbb{R}^{n \times m} \to \mathbb{R}$ is represented as $f(X) = \widetilde{f}(\sigma)$, where $n \geq m$ and $\sigma = (\sigma_1, \cdots, \sigma_m)$ denotes the singular values of X. If f is* *twice weakly differentiable, then its Laplacian is expressed as*

$$\Delta f = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\partial^2 f}{\partial X_{ij}^2} = 2 \sum_{i<j} \frac{\sigma_i \frac{\partial \widetilde{f}}{\partial \sigma_i} - \sigma_j \frac{\partial \widetilde{f}}{\partial \sigma_j}}{\sigma_i^2 - \sigma_j^2} + (n-m) \sum_{i=1}^{m} \frac{1}{\sigma_i} \frac{\partial \widetilde{f}}{\partial \sigma_i} + \sum_{i=1}^{m} \frac{\partial^2 \widetilde{f}}{\partial \sigma_i^2}. \quad (14)$$

4.1 Fundamental results

First, we consider estimators of the form (3) with $\phi_i(\sigma) = r_i/\sigma_i^2$, where each r_i is constant. This class includes the maximum likelihood estimator ($r_i = 0$) and the Efron–Morris estimator (4) ($r_i = n - m - 1$). The unbiased loss estimator (13) is

$$\lambda^{\mathrm{U}}(X) = nm + \sum_{i=1}^{m} \frac{r_i(r_i - 2(n-m-1))}{\sigma_i(X)^2} - 4 \sum_{i<j} \frac{r_i - r_j}{\sigma_i(X)^2 - \sigma_j(X)^2}. \quad (15)$$

4.1.1 Downward correction

For a normal mean vector, Johnstone (1988) showed that the unbiased loss estimator for the maximum likelihood estimator is improved by downward correction (Lemma 2). The following Theorem gives a class of matrix mean estimators for which downward correction provides improvement, and which includes the maximum likelihood estimator.

Theorem 2. *Consider an estimator $\hat{M}$ of the form* (3) *with $\phi_i(\sigma) = r_i/\sigma_i^2$. If $r_i + i$ is non-decreasing in i and $n - m - 2i - 1 - 2r_i > 0$ for every i, then the loss estimator*

$$\lambda(X) = \lambda^{\mathrm{U}}(X) - \sum_{i=1}^{m} c_i \sigma_i(X)^{-2}, \quad c_i = \frac{4}{m}(n - m - 2i - 1 - 2r_i),$$

dominates the unbiased loss estimator $\lambda^{\mathrm{U}}(X)$ in (15).

Proof. From the assumptions on r_i, we have $c_1 \geq \cdots \geq c_m > 0$. Therefore,

$$\begin{aligned}
\sum_{i<j} \frac{c_i \sigma_i^{-2} - c_j \sigma_j^{-2}}{\sigma_i^2 - \sigma_j^2} &= \sum_{i<j} \frac{(c_i - c_j)\sigma_j^2 + c_j(\sigma_j^2 - \sigma_i^2)}{\sigma_i^2 \sigma_j^2 (\sigma_i^2 - \sigma_j^2)} \\
&= \sum_{i<j} \frac{c_i - c_j}{\sigma_i^2(\sigma_i^2 - \sigma_j^2)} - \sum_{i<j} \frac{c_j}{\sigma_i^2 \sigma_j^2} \\
&\geq - \sum_{i<j} \frac{c_j}{\sigma_i^2 \sigma_j^2} \\
&\geq - \sum_{i<j} \frac{c_j}{\sigma_j^4} \\
&= - \sum_{i=1}^{m} (i-1) \frac{c_i}{\sigma_i^4}. \quad (16)
\end{aligned}$$

Let $h(X) = -\sum_{i=1}^m c_i \sigma_i(X)^{-2}$ and $g(X) = -U\Sigma\Phi(\Sigma)V^\top$. From (14) and (16), we have

$$\begin{aligned}\Delta h &= 4\sum_{i<j} \frac{c_i\sigma_i^{-2} - c_j\sigma_j^{-2}}{\sigma_i^2 - \sigma_j^2} + 2(n-m-3)\sum_{i=1}^m c_i\sigma_i^{-4} \\ &\geq \sum_{i=1}^m 2(n-m-2i-1)c_i\sigma_i^{-4}.\end{aligned}$$

Also, we have

$$\begin{aligned}g^\top \nabla h &= \frac{\mathrm{d}}{\mathrm{d}t}\, h(X+tg)|_{t=0} \\ &= -\frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^m c_i \left(\sigma_i - \frac{r_i}{\sigma_i}t\right)^{-2}\Bigg|_{t=0} \\ &= -2\sum_{i=1}^m r_i c_i \sigma_i^{-4}.\end{aligned}$$

Therefore,

$$\begin{aligned}-2\Delta h - 4g^\top\nabla h + h^2 &\leq \sum_{i=1}^m (c_i^2 - 4(n-m-2i-1-2r_i)c_i)\sigma_i^{-4} + 2\sum_{i<j} c_ic_j\sigma_i^{-2}\sigma_j^{-2} \\ &= -(m-1)\sum_{i=1}^m c_i^2\sigma_i^{-4} + 2\sum_{i<j} c_ic_j\sigma_i^{-2}\sigma_j^{-2} \\ &= -\sum_{i<j}(c_i\sigma_i^{-2} - c_j\sigma_j^{-2})^2 \\ &\leq 0.\end{aligned}$$

Hence, from Lemma 1, $\lambda(X) = \lambda^{\mathrm{U}}(X) + h(X)$ dominates $\lambda^{\mathrm{U}}(X)$. □

Corollary 2. *Suppose $n - 3m - 1 > 0$. For the maximum likelihood estimator $\hat{M}(X) = X$, the loss estimator*

$$\lambda(X) = nm - \sum_{i=1}^m c_i\sigma_i(X)^{-2}, \quad c_i = \frac{4}{m}(n-m-2i-1), \tag{17}$$

dominates the unbiased loss estimator $\lambda^{\mathrm{U}}(X) = nm$.

Proof. It is obvious from Theorem 2 and $r_i = 0$. □

4.1.2 Upward correction

For a normal mean vector, Johnstone (1988) showed that the unbiased loss estimator for the James–Stein estimator is improved by upward correction (Lemma 3). The following Theorem gives a class of matrix mean estimators for which upward correction provides improvement, and which includes the Efron–Morris estimator (4).

Theorem 3. *Consider an estimator $\hat{M}$ of the form* (3) *with $\phi_i(\sigma) = r_i/\sigma_i^2$. If r_i is non-decreasing in i and $2r_i - n + m + 3 > 0$ for every i, then the loss estimator*

$$\lambda(X) = \lambda^{\mathrm{U}}(x) + \sum_{i=1}^m c_i \sigma_i(X)^{-2}, \quad c_i = \frac{4}{m}(2r_i - n + m + 3),$$

dominates the unbiased loss estimator $\lambda^{\mathrm{U}}(X)$ in (15).

Proof. From the assumptions on r_i, we have $0 < c_1 \leq \cdots \leq c_m$. Therefore,

$$\begin{aligned} \sum_{i<j} \frac{c_i \sigma_i^{-2} - c_j \sigma_j^{-2}}{\sigma_i^2 - \sigma_j^2} &= \sum_{i<j} \frac{c_i(\sigma_j^2 - \sigma_i^2) + \sigma_i^2 (c_i - c_j)}{\sigma_i^2 \sigma_j^2 (\sigma_i^2 - \sigma_j^2)} \\ &= -\sum_{i<j} \frac{c_i}{\sigma_i^2(\sigma_i^2 - \sigma_j^2)} + \sum_{i<j} \frac{c_i - c_j}{\sigma_j^2} \\ &\leq 0. \end{aligned} \tag{18}$$

Let $h(X) = \sum_{i=1}^m c_i \sigma_i(X)^{-2}$ and $g(X) = -U\Sigma\Phi(\Sigma)V^\top$. From (14) and (18), we have

$$\begin{aligned} \Delta h &= -4 \sum_{i<j} \frac{c_i \sigma_i^{-2} - c_j \sigma_j^{-2}}{\sigma_i^2 - \sigma_j^2} - 2(n-m-3) \sum_{i=1}^m c_i \sigma_i^{-4} \\ &\geq -\sum_{i=1}^m 2(n-m-3) c_i \sigma_i^{-4}. \end{aligned}$$

Also, we have

$$\begin{aligned} g^\top \nabla h &= \frac{\mathrm{d}}{\mathrm{d}t} \, h(X + tg)|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \sum_{i=1}^m c_i \left(\sigma_i - \frac{r_i}{\sigma_i} t \right)^{-2} \Bigg|_{t=0} \\ &= 2 \sum_{i=1}^m r_i c_i \sigma_i^{-4}. \end{aligned}$$

Therefore,

$$\begin{aligned}
-2\Delta h - 4g^\top \nabla h + h^2 &\le \sum_{i=1}^m (c_i^2 - 4(2r_i - n + m + 3)c_i)\sigma_i^{-4} + 2\sum_{i<j} c_i c_j \sigma_i^{-2}\sigma_j^{-2} \\
&= -(m-1)\sum_{i=1}^m c_i^2 \sigma_i^{-4} + 2\sum_{i<j} c_i c_j \sigma_i^{-2}\sigma_j^{-2} \\
&= -\sum_{i<j} (c_i\sigma_i^{-2} - c_j\sigma_j^{-2})^2 \\
&\le 0.
\end{aligned}$$

Hence, from Lemma 1, $\lambda(X) = \lambda^{\mathrm{U}}(X) + h(X)$ dominates $\lambda^{\mathrm{U}}(X)$. □

Corollary 3. *Suppose $n - m - 1 > 0$. For the Efron–Morris estimator $\hat{M}_{\mathrm{EM}}$ in (4), the loss estimator*

$$\lambda(X) = nm - (n-m-1)^2 \sum_{i=1}^m \sigma_i(X)^{-2} + c\sum_{i=1}^m \sigma_i(X)^{-2}, \quad c = \frac{4}{m}(n-m+1), \quad (19)$$

dominates the unbiased loss estimator $\lambda^{\mathrm{U}}(X) = nm - (n-m-1)^2 \sum_{i=1}^m \sigma_i(X)^{-2}$.

Proof. It is obvious from Theorem 3 and $r_i = n - m - 1$. □

4.1.3 Border between downward and upward correction

Here, we consider estimators of the form (3) with $\phi_i(\sigma) = r/\sigma_i^2$. The maximum likelihood estimator ($r = 0$) and the Efron–Morris estimator ($r = n - m - 1$) belong to this class. The following Theorem provides the border between downward and upward correction.

Theorem 4. *Consider an estimator $\hat{M}$ of the form (3) with $\phi_i(\sigma) = r/\sigma_i^2$.*

(i) If $0 \le r < (n - 2m - 2)/2$, then the loss estimator with downward correction

$$\lambda(X) = \lambda^{\mathrm{U}}(X) - c\sum_{i=1}^m \sigma_i(X)^{-2}, \quad 0 \le c \le \frac{4}{m}(n - 2m - 2r - 2),$$

dominates the unbiased loss estimator $\lambda^{\mathrm{U}}(X) = nm + r(r - 2(n-m-1)) \sum_{i=1}^m \sigma_i(X)^{-2}$.

(ii) If $r > (n - 2m - 2)/2$, then the loss estimator with upward correction

$$\lambda(X) = \lambda^{\mathrm{U}}(X) + c\sum_{i=1}^m \sigma_i(X)^{-2}, \quad 0 \le c \le \frac{4}{m}(-n + 2m + 2r + 2),$$

dominates the unbiased loss estimator $\lambda^{\mathrm{U}}(X) = nm + r(r - 2(n-m-1)) \sum_{i=1}^m \sigma_i(X)^{-2}$.

Proof. Let $h(X) = c\sum_{i=1}^m \sigma_i(X)^{-2}$ and $g(X) = -U\Sigma\Phi(\Sigma)V^\top$. From (14), we have

$$\Delta h = 4c\sum_{i<j}\sigma_i^{-2}\sigma_j^{-2} - 2(n-m-3)c\sum_{i=1}^m \sigma_i^{-4}.$$

Also, we have

$$\begin{aligned} g^\top \nabla h &= \frac{\mathrm{d}}{\mathrm{d}t}\, h(X+tg)|_{t=0} \\ &= c\frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^m \left(\sigma_i - \frac{r}{\sigma_i}t\right)^{-2}\Bigg|_{t=0} \\ &= 2cr\sum_{i=1}^m \sigma_i^{-4}. \end{aligned}$$

Therefore,

$$\begin{aligned} -2\Delta h - 4g^\top\nabla h + h^2 &= (c^2 + 4(n-m-2r-3)c)\sum_{i=1}^m \sigma_i^{-4} + (2c^2 - 8c)\sum_{i<j}\sigma_i^{-2}\sigma_j^{-2} \\ &\leq (mc^2 + 4(n-2m-2r-2)c)\sum_{i=1}^m \sigma_i(X)^{-4}, \end{aligned}$$

where we used

$$(m-1)\sum_{i=1}^m \sigma_i^{-4} - 2\sum_{i<j}\sigma_i^{-2}\sigma_j^{-2} = \sum_{i<j}(\sigma_i^{-2} - \sigma_j^{-2})^2 \geq 0.$$

Hence, from Lemma 1, $\lambda(X) = \lambda^{\mathrm{U}}(X) + h(X)$ dominates $\lambda^{\mathrm{U}}(X)$ if

$$(mc^2 + 4(n-2m-2r-2)c)\sum_{i=1}^m \sigma_i(X)^{-4} \leq 0,$$

which holds when c is between 0 and $-(4/m)(n-2m-2r-2)$. □

4.2 Estimators with scalar shrinkage

The modified Efron–Morris estimator (5) adds scalar shrinkage to the Efron–Morris estimator (4). Tsukuma (2008) provided a general condition on estimators of the form (3) to be improved by additional scalar shrinkage. Here, we improve on the unbiased loss estimators for matrix mean estimators of the form (3) with scalar shrinkage:

$$\phi_i(\sigma) = \frac{r_i}{\sigma_i^2} + \frac{s}{\sum_{i=1}^m \sigma_i^2},$$

where $s > 0$.

Theorem 5. *Let $\hat{M}_0$ and $\hat{M}$ be estimators of the form* (3) *with $\phi_i(\sigma) = r/\sigma_i^2$ and $\phi_i(\sigma) = r_i/\sigma_i^2 + s/(\sum_{i=1}^m \sigma_i^2)$, respectively. If the loss estimator*

$$\overline{\lambda}(X) = \overline{\lambda}^{\mathrm{U}}(X) + \sum_{i=1}^m c_i \sigma_i(X)^{-2}$$

dominates the unbiased loss estimator $\overline{\lambda}^{\mathrm{U}}(X)$ for $\hat{M}_0$, then the loss estimator

$$\lambda(X) = \lambda^{\mathrm{U}}(X) + \sum_{i=1}^m c_i \sigma_i(X)^{-2}$$

dominates the unbiased loss estimator $\lambda^{\mathrm{U}}(X)$ for $\hat{M}$.

Proof. Let $h(X) = \sum_{i=1}^m c_i \sigma_i(X)^{-2}$, $g_0(X) = \hat{M}_0(X) - X$ and $g(X) = \hat{M}(X) - X$. From Lemma 1, the risk difference between the loss estimator $\overline{\lambda}(X) = \overline{\lambda}^{\mathrm{U}}(X) + h(X)$ and the unbiased loss estimator $\overline{\lambda}^{\mathrm{U}}(X)$ for $\hat{M}_0$ is expressed as

$$\mathrm{E}_M[L^*(M, \hat{M}_0(X), \overline{\lambda}(X))] - \mathrm{E}_M[L^*(M, \hat{M}_0(X), \overline{\lambda}^{\mathrm{U}}(X))] = \mathrm{E}_M[-2\Delta h - 4 g_0^\top \nabla h + h^2].$$

Since $\overline{\lambda}(X)$ dominates $\overline{\lambda}^{\mathrm{U}}(X)$, we have

$$\mathrm{E}_M[-2\Delta h - 4 g_0^\top \nabla h + h^2] \leq 0$$

for every M. Also, from $s > 0$, we obtain

$$\begin{aligned} g^\top \nabla h &= \frac{\mathrm{d}}{\mathrm{d}t} \, h(X + tg)|_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \sum_{i=1}^m c_i \left(\sigma_i - \frac{r_i}{\sigma_i} t - \frac{s \sigma_i}{\sum_j \sigma_j^2} t \right)^{-2} \Bigg|_{t=0} \\ &= g_0^\top \nabla h + 2s \sum_{i=1}^m c_i \sigma_i^{-2} \left(\sum_{i=1}^m \sigma_i^2 \right)^{-1} \\ &\geq g_0^\top \nabla h. \end{aligned}$$

Therefore, the risk difference between the loss estimator $\lambda(X) = \lambda^{\mathrm{U}}(X) + h(X)$ and the unbiased loss estimator $\lambda^{\mathrm{U}}(X)$ for $\hat{M}$ is

$$\begin{aligned} \mathrm{E}_M[L^*(M, \hat{M}(X), \lambda(X))] - \mathrm{E}_M[L^*(M, \hat{M}(X), \lambda^{\mathrm{U}}(X))] &= \mathrm{E}_M[-2\Delta h - 4 g^\top \nabla h + h^2] \\ &\leq \mathrm{E}_M[-2\Delta h - 4 g_0^\top \nabla h + h^2] \\ &\leq 0 \end{aligned}$$

for every M. □

Corollary 4. *Suppose $n-m-1>0$. For the Efron–Morris estimator with scalar shrinkage $\hat{M}_{\rm MEM}$ in* (5)*, the loss estimator*

$$\lambda(X) = \lambda^{\rm U}(X) + c\sum_{i=1}^m c_i \sigma_i(X)^{-2}, \quad c = \frac{4}{m}(n-m+1), \tag{20}$$

dominates the unbiased loss estimator $\lambda^{\rm U}(X)$.

Proof. It is obvious from Theorem 5 and Corollary 3. □

4.3 Reduced-rank estimators

Finally, we consider the reduced-rank estimator with rank k, which is represented as (3) with (7). Recently, Mukherjee et al. (2015) derived the degrees of freedom for reduced-rank regression. From their results, the unbiased loss estimator is obtained as

$$\lambda^{\rm U}(X) = \sum_{i=k+1}^m \sigma_i(X)^2 - nm + 2\left(\sum_{i=1}^k \sum_{j=k+1}^m \frac{\sigma_i(X)^2+\sigma_j(X)^2}{\sigma_i(X)^2-\sigma_j(X)^2} + nk\right). \tag{21}$$

Theorem 6. *Consider the reduced-rank estimator with rank k. If $n-m-2k-1 \geq 0$, then the loss estimator*

$$\lambda(X) = \lambda^{\rm U}(X) - \sum_{i=1}^k c_i \sigma_i(X)^{-2}, \quad c_i = \frac{4}{k}(n-m-2i-1), \tag{22}$$

dominates the unbiased loss estimator $\lambda^{\rm U}(X)$ in (21)*.*

Proof. Let $h(X) = -\sum_{i=1}^k c_i \sigma_i(X)^{-2}$ and $g(X) = -U\Sigma\Phi(\Sigma)V^\top$, where $\phi_i(\sigma)$ is defined as (7). From (14) and (16), we have

$$\begin{aligned}\Delta h &= 4\sum_{i=1}^k \sum_{j=i+1}^k \frac{c_i\sigma_i^{-2} - c_j\sigma_j^{-2}}{\sigma_i^2-\sigma_j^2} + 4\sum_{i=1}^k\sum_{j=k+1}^m \frac{c_i\sigma_i^{-2}}{\sigma_i^2-\sigma_j^2} + 2(n-m-3)\sum_{i=1}^k c_i\sigma_i^{-4} \\ &\geq -4\sum_{i=1}^k (i-1)c_i\sigma_i^{-4} + 2(n-m-3)\sum_{i=1}^k c_i\sigma_i^{-4} \\ &= \sum_{i=1}^k 2(n-m-2i-1)c_i\sigma_i^{-4}.\end{aligned}$$

Also, we have

$$g^\top \nabla h = \frac{\rm d}{{\rm d}t}\, h(X+tg)|_{t=0} = 0,$$

since $\sigma_i(X+tg)=\sigma_i(X)$ for $1 \leq i \leq k$. Therefore,

$$
\begin{aligned}
-2\Delta h - 4g^\top \nabla h + h^2 &\leq \sum_{i=1}^{k} (c_i^2 - 4(n-m-2i-1)c_i)\sigma_i^{-4} + 2\sum_{i<j} c_i c_j \sigma_i^{-2}\sigma_j^{-2} \\
&= -(k-1)\sum_{i=1}^{k} c_i^2 \sigma_i^{-4} + 2\sum_{i=1}^{k}\sum_{j=i+1}^{k} c_i c_j \sigma_i^{-2}\sigma_j^{-2} \\
&= -\sum_{i=1}^{k}\sum_{j=i+1}^{k} (c_i\sigma_i^{-2} - c_j\sigma_j^{-2})^2 \\
&\leq 0.
\end{aligned}
$$

Hence, from Lemma 1, $\lambda(X) = \lambda^{\mathrm{U}}(X) + h(X)$ dominates $\lambda^{\mathrm{U}}(X)$. □

5 Numerical results

In this section, we show the performance of improved loss estimators by numerical experiments. The risk functions of loss estimators are computed by the Monte Carlo method with sample size 10^6. For each loss estimator $\lambda(X)$, we present the percentage improvement in risk over the unbiased loss estimator $\lambda^{\mathrm{U}}(X)$:

$$
100\frac{\mathrm{E}_M[L^*(M,\hat{M}(X),\lambda^{\mathrm{U}}(X))] - \mathrm{E}_M[L^*(M,\hat{M}(X),\lambda(X))]}{\mathrm{E}_M[L^*(M,\hat{M}(X),\lambda^{\mathrm{U}}(X))]}. \tag{23}
$$

In the results below, we plot (23) as a function of the singular values of the mean matrix $\sigma_1 = \sigma_1(M), \dots, \sigma_m = \sigma_m(M)$.

5.1 Maximum likelihood estimator

Figure 1 shows the percentage improvement in risk of loss estimators for the maximum likelihood estimator $\hat{M}(X) = M$. The unbiased loss estimator is $\lambda^{\mathrm{U}}(X) = nm$. Our loss estimator (17) and the improved loss estimator by Johnstone (1988) $\lambda^{\mathrm{J}}(X) = nm - 2(nm-4)/\|X\|_{\mathrm{F}}^2$ are compared. Note that we formally extended Johnstone's loss estimator to matrix mean case by considering the vectorization of M and X.

Figure 1 (a) shows the case $n=8$, $m=2$ and $\sigma_1 = 10$. Our loss estimator performs better than the unbiased loss estimator and the risk reduction increases as σ_2 decreases, whereas Johnstone's loss estimator has almost the same risk with the unbiased loss estimator. Figure 1 (b) shows the case $n=8$, $m=2$ and $\sigma_2 = 0$. Though Johnstone's loss estimator performs best when σ_1 is small, its risk becomes almost the same as that of the unbiased estimator as σ_1 increases. On the other hand, our loss estimator has constant risk reduction even when σ_1 is large.

In summary, our loss estimator works well when several singular values are close to zero. Thus, it is effective when the mean matrix has low rank. These properties are similar to the Frobenius risk of the Efron–Morris estimator (4) and also the Bayes estimator with respect to the singular value shrinkage prior (6).

(a) (b)

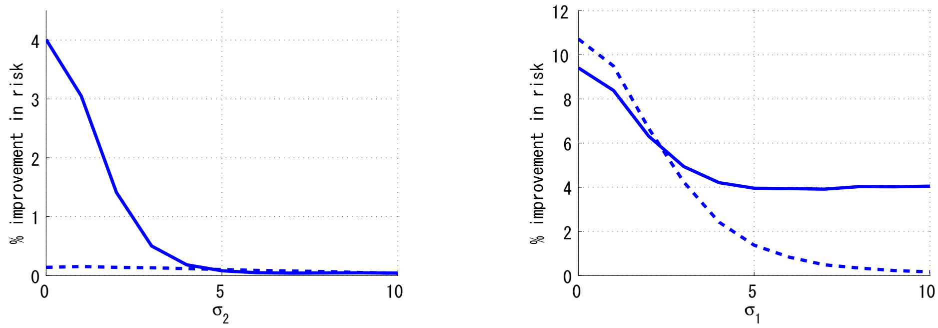

Figure 1: Percentage improvement in risk of our loss estimator (17) (solid line) and Johnstone's loss estimator (dashed line) over the unbiased loss estimator for the maximum likelihood estimator when $n = 8$ and $m = 2$. (a) $\sigma_1 = 10$. (b) $\sigma_2 = 0$.

5.2 Efron–Morris estimator

Figure 2 shows the percentage improvement in risk of our loss estimator (19) for the Efron–Morris estimator (4) when $n = 6$ and $m = 2$. Our loss estimator performs better than the unbiased loss estimator and the risk reduction is large when several singular values are close to zero. In particular, the percentage improvement in risk is almost constant when the mean matrix has low rank, as shown in Figure 2 (b).

5.3 Efron–Morris estimator with scalar shrinkage

Figure 3 shows the percentage improvement in risk of our loss estimator (20) for the Efron–Morris estimator with scalar shrinkage (5) when $n = 6$ and $m = 2$. Similarly to the Efron-Morris estimator without scalar shrinkage (Figure 2), our loss estimator has large risk reduction when several singular values are close to zero and the percentage improvement in risk is almost constant when the mean matrix has low rank. The percentage improvement in risk is larger than that for the Efron-Morris estimator without shrinkage.

(a) (b)

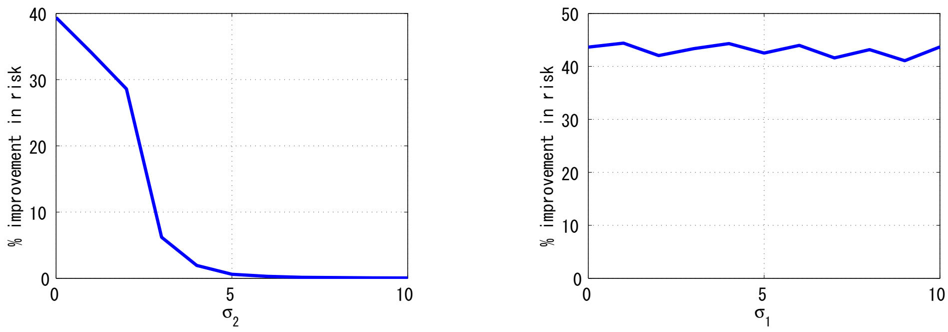

Figure 2: Percentage improvement in risk of our loss estimator (19) over the unbiased loss estimator for the Efron–Morris estimator when $n = 6$ and $m = 2$. (a) $\sigma_1 = 10$ (b) $\sigma_2 = 0$.

(a) (b)

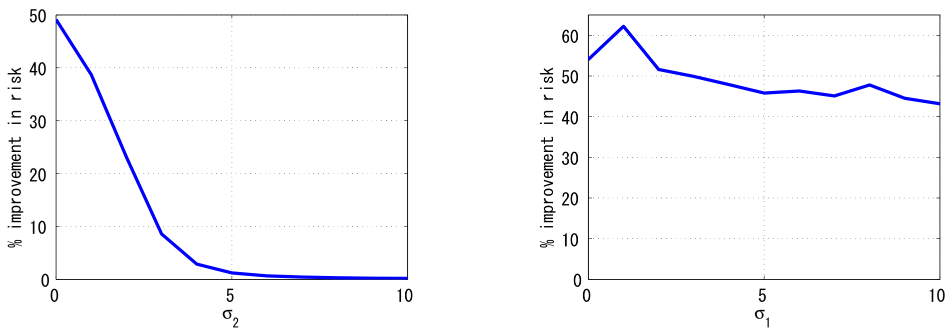

Figure 3: Percentage improvement in risk of our loss estimator (20) over the unbiased loss estimator for the Efron–Morris estimator with scalar shrinkage (5) when $n = 6$ and $m = 2$. (a) $\sigma_1 = 10$. (b) $\sigma_2 = 0$.

5.4 Reduced-rank estimator

Figure 4 shows the percentage improvement in risk of our loss estimator (22) for the reduced-rank estimator (7). Compared to Figure 1-3, the percentage improvement in risk is not large. Figure 4 (a) shows the case $n = 6$, $m = 2$, $k = 1$, and $\sigma_2 = 0$ and Figure 4 (b) shows the case $n = 9$, $m = 3$, $k = 2$, and $\sigma_2 = \sigma_3 = 0$. In both cases, the risk reduction by our loss estimator increases as σ_1 goes to zero.

(a) (b)

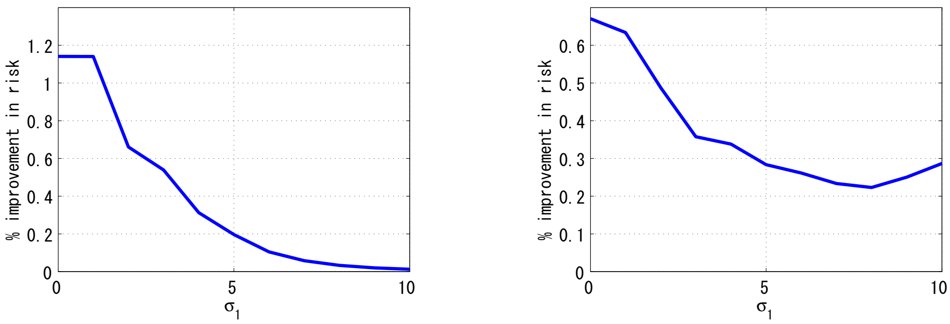

Figure 4: Percentage improvement in risk of our loss estimator (22) over the unbiased loss estimator for the reduced-rank estimator (7). (a) $n = 6$, $m = 2$, $k = 1$, and $\sigma_2 = 0$. (b) $n = 9$, $m = 3$, $k = 2$, and $\sigma_2 = \sigma_3 = 0$.

6 Concluding remarks

In this study, we investigated loss estimation for a normal mean matrix. We developed loss estimators that dominate the unbiased loss estimator for a broad class of matrix mean estimators, including the Efron–Morris estimator with and without scalar shrinkage. Our results include an extension of the results of Johnstone (1988) for a normal mean vector. We also provided improved loss estimators for reduced-rank estimators. We confirmed the effectiveness of improved loss estimators by numerical results.

In practice, it is unreasonable to use a negative value as an estimate of loss. Therefore, if a loss estimator $\lambda(X)$ takes negative values with nonzero probability, its positive-part version $\max(0, \lambda(X))$ dominates the original $\lambda(X)$. For example, the unbiased loss estimator and also our improved estimator for the Efron–Morris estimator take negative values. However, numerical experiments showed that the positive-part version of the unbiased loss estimator is not dominated by the positive-part version of our improved loss estimator. It is a future problem to develop loss estimators that improve upon the positive-part version of the unbiased loss estimator in such case.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 14J09148 and a grant from the Simons Foundation (#209035 and #418098 to William Strawderman).

References

Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika* **68**, 265–74.

Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika* **1**, 211–218.

Efron, B. & Morris, C. (1972). Empirical Bayes on vector observations: an extension of Stein's method. *Biometrika* **59**, 335–347.

Efron, B. & Morris, C. (1976). Multivariate empirical Bayes and estimation of covariance matrices. *The Annals of Statistics* **4**, 22–32.

Fourdrinier, D. & Wells, M. T. (2012). On improved loss estimation for shrinkage estimators. *Statistical Science* **27**, 61–81.

Fourdrinier, D. & Strawderman, W. E. (2003). Bayes and unbiased estimators of loss. *Annals of the Institute of Statistical Mathematics* **55**, 803–816.

James, A. T. (1954). Normal multivariate analysis and the orthogonal group. *The Annals of Mathematical Statistics* **25**, 40–75.

Johnstone, I. M. (1988). On inadmissibility of some unbiased estimators of loss. *Statistical Decision Theory and Related Topics* **4**, 361–379.

Lu, K. L. & Berger, J. O. (1989). Estimation of normal means: frequentist estimation of loss. *The Annals of Statistics* **17**, 890–906.

Matsuda, T. & Komaki, F. (2015). Singular value shrinkage priors for Bayesian prediction. *Biometrika* **102**, 843–854.

Mukherjee, A., Chen, K., Wang, N. & Zhu, J. (2015). On the degrees of freedom of reduced-rank estimators in multivariate regression. *Biometrika* **102**, 457–477.

Narayanan, R. & Wells, M. T. (2015). Improved loss estimation for the Lasso: a variable selection tool. *Sankhya B* **77**, 45–74.

Reinsel, G. C. & Velu, R. P. (1998). *Multivariate Reduced-Rank Regression.* New York: Springer-Verlag.

Sheena, Y. (1995). Unbiased estimator of risk for an orthogonally invariant estimator of a covariance matrix. *Journal of the Japan Statistical Society* **25**, 35–48.

Stein, C. (1974). Estimation of the mean of a multivariate normal distribution. *Proceedings of Prague Symposium on Asymptotic Statistics* **2**, 345–381.

Takemura, A. & Kuriki, S. (1999). Shrinkage to smooth non-convex cone: principal component analysis as Stein estimation. *Communications in Statistics, Theory and Methods* **28**, 651–669.

Tsukuma, H. (2008). Admissibility and minimaxity of Bayes estimators for a normal mean matrix. *Journal of Multivariate Analysis* **99**, 2251–2264.

Zheng, Z. (1986). On estimation of matrix of normal mean. *Journal of Multivariate Analysis* **18**, 70–82.